NPRDC TR 86-26

JULY 1986

COLLECTIVE TRAINING STANDARDS DEVELOPMENT: PROBLEM ANALYSIS

NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER
San Diego, California 92152

NPRDC TR 86-26 July 1986

COLLECTIVE TRAINING STANDARDS DEVELOPMENT: PROBLEM ANALYSIS

Sandra K. Wetzel-Smith
Steve R. Mitchell

Reviewed by
J. McGrath

Approved by
J. McLachlan

Released by
B. E. Bacon
Captain, U.S. Navy
Commanding Officer

Navy Personnel Research and Development Center
San Diego, California 92152-6800

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

A169 457

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NPRDC TR 86-26	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Navy Personnel Research and Development Center 6b. OFFICE SYMBOL (If applicable) Code 52	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) San Diego, CA 92152-6800	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Headquarters, U.S. Marine Corps 8b. OFFICE SYMBOL (If applicable) TDC-20	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Washington, DC 20380-0001	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
PE62763			

11. TITLE (Include Security Classification)

COLLECTIVE TRAINING STANDARDS DEVELOPMENT: PROBLEM ANALYSIS

12. PERSONAL AUTHOR(S)
Wetzel-Smith, S. K.; Mitchell, S. R.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim Report	FROM 84 Dec TO 85 Dec	1986 July	23

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
05	09	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
~~Collective training standards~~, team training standards, team performance standards, assessing unit performance, collective performance standards.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The Marine Corps requires valid, reliable, and cost-effective measures for developing collective training standards (CTSs). This report identifies and describes procedures now in use in the Armed Forces for developing CTSs and reviews the literature regarding unit performance and measurement from 1960 to 1984. It was concluded that a functional analysis of the organizational responsibilities of the unit is required to support the development of CTSs. Recommendations address methods for developing and validating collective training standards.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL Sandra K. Wetzel-Smith	22b. TELEPHONE (Include Area Code) (619) 225-6539	22c. OFFICE SYMBOL Code 52

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

A

FOREWORD

This research was performed in response to a request from Headquarters, U.S. Marine Corps to establish a methodology for the development of collective training standards (CTSs). The program element number was PE 62763, under the sponsorship of the Deputy Chief of Staff for Training, code TDC-20.

This effort was conducted to identify the existing procedures for developing CTSs and to determine valid and cost-effective methods for developing CTSs for use in the Marine Corps into the mid-1990s.

B. E. BACON
Captain, U.S. Navy
Commanding Officer

J. W. TWEEDDALE
Technical Director

SUMMARY

Problem

The Marine Corps' need for methods to develop both collective training standards (CTSs) and measures of unit readiness requires the development of valid, reliable, and cost-effective methods for defining unit performance requirements.

Objective

The objectives of this effort were to identify the existing procedures for developing CTSs and to determine valid and cost-effective methods for developing CTSs for use in the Marine Corps into the mid-1990s.

Approach

Existing performance requirements and standards development procedures were reviewed for adequacy, applicability, and usefulness to the Marine Corps training and evaluation needs of the mid-1990s.

Research on methodological issues in the development of unit/team performance standards was reviewed and discussed.

Findings and Discussion

1. Development of valid CTSs requires: (a) a definition of team or unit composition, (b) a baseline set of behaviors of the unit's mission responsibilities, (c) descriptions of the performance conditions, and (d) a description of the desired outcomes of the unit's behavior.

2. Scientific findings on training issues relevant to the CTS methods are: (a) Collective task performance depends on individual subtask proficiency; (b) when the tasks require more direct coordination, the unit skill levels are higher than the level of individual skills; (c) units that receive feedback about their performance improve more than those who do not receive feedback.

3. The major methodological problem in developing unit training standards is the development of criterion variables that are objective, recordable, reliable, and discriminate between levels of performance. The apparent solution to this problem during the time period of interest is to use subject matter experts (SMEs) more effectively for analyzing the tasks under scrutiny.

4. Some of the alternative methodologies for developing CTSs are to use experienced officer judgements in combat scenarios, DELPHI-related methods, Marine Corps Combat Readiness Evaluation System (MCCRES) mission performance standards (MPS) as a basis for developing unit performance standards, and the Army's engagement simulations (ESs) for specification of unit performance measures.

5. The Army and Navy now use the following useful common procedures for validating CTSs. They (a) begin with a formal description of the unit's mission, (b) monitor the unit's performance during the mission, and (c) compare the unit's performance with the established performance standards. All the validations consider the progression from individual to unit training to be of key importance.

6. To develop an evaluation system that is capable of producing valid judgements of unit performance requires a functional analysis of the organizational responsibilities and a context for assigning mission/task responsibilities within a unit, between coordinated units, and by echelon.

7. The methods currently used to establish unit performance standards range from quasi-empirical to highly subjective techniques. The cost of the techniques vary with objectivity--the quasi-empirical being the most expensive and the subjective techniques being the least expensive. Among the research community, confidence in the validity of standards seemed highest with quasi-empirical methods.

Recommendations

1. Develop a method for determining unit functional responsibilities within mission tasks. The resulting data would serve as a basis for determining collective training requirements and their supporting individual training requirements.

2. Develop a method for identifying and formatting effective standards for guiding the training and evaluation of collective training requirements.

3. Investigate the feasibility of the use of the DELPHI approach in obtaining a consensus of SME opinion for the establishment of collective training standards.

4. Develop cost-effective approaches for validating the effectiveness of collective training standards.

CONTENTS

INTRODUCTION

Problem

The Marine Corps maintains a high level of readiness to respond to a range of situations. Individual Marines must train to perform the critical skills that contribute to readiness. The outcome of combat depends on the performance of individual Marines on individual as well as performance on collective tasks. Therefore, individual skills should be integrated into team skills to meet mission responsibilities.

Nearly every existing organization has devised ways to evaluate its own performance. Many methods guided the development of evaluation systems. Yet, nearly every organization that has one wants a better one--if not now, at least in the future. The most common complaint about evaluation systems is that they do not tell you what you need to know: Exactly why an individual or unit succeeded or failed.

Measuring readiness requires the development of a valid, reliable, and cost-effective method for defining team performance. When team performance requirements are defined, training and evaluation standards can be developed. These standards will accurately assess the level of readiness and direct the diagnosis for remedial or additional training. The Marine Corps needs methods for developing both collective training standards (CTSs) and readiness measures. Existing requirements and methodologies must be reviewed to determine their adequacy, applicability, and usefulness to the Marine Corps training and evaluation requirements of the 1990s.

Objective

The objectives of this effort were to identify the existing procedures for developing CTSs and to determine valid and cost-effective methods for developing CTSs for use in the Marine Corps into the mid-1990s.

Background

Any combat performance evaluation system must be based on valid performance standards. These standards must represent the most important elements of behavior in battle within the context of the organization of the largest unit committed to the conflict. Combat effectiveness is a multidimensional phenomenon that is not yet reflected adequately in any single objective measurement or number index. This is particularly true when the performance to be evaluated includes the complex interactions between units necessary to ensure mission success.

Combat evaluation systems tend to be event oriented (i.e., the order was given within 30 minutes or the troops proceeded in a wedge formation). Under these systems, qualitative information about the appropriateness of the performance/decision/command goes largely unmeasured. A more serious failing of current evaluation standards is that they do not measure the effectiveness of team interactions performed in integrated maneuvers. Rarely is there an explicit measure to judge whether Team A completed its mission in a way that allowed Team B to accomplish its mission. This is particularly true of judgments about the quality of performance between levels of command.

Current evaluation systems do not allow determination of cause and effect relationships between intermediate actions and final mission outcomes. Current measures of unit performance do not provide a useful summary of exercise events with valid estimates of both readiness and guidance for training. Using these types of standards, it is difficult to diagnose specific problems in unit training and identify the critical tasks to be evaluated, how performance should be measured, and the coordinated relationships within and between units.

APPROACH

Research issues in the development of performance standards were investigated and are discussed. Examination of each methodology focused on methods for developing unit performance standards and emphasized (1) evaluation, organization, and training issues, (2) analyses of methods for identifying and validating standards, and (3) developing new collective performance standards.

FINDINGS

The importance and difficulty of developing methods for evaluating team performance are widely acknowledged (Wagner, Hibbits, Rosenblatz, & Schultz, 1977). As early as 1955, Wagner et al. (1977) cited the absence of satisfactory proficiency measures. In 1962, Glaser reported that, although the importance of developing proficiency measures for team performance was frequently considered, little had been accomplished. More than 10 years later, Obermayer (1974) concluded that a means of objectively measuring skills in team settings was an elusive goal, and the Defense Science Board (1975) called team performance measurement a "fundamental stumbling block to progress" in improving team performance.

Evaluation Issues

The complex process of evaluating team performance requires understanding the fundamental conditions under which performance occurs. The first issue is how to define the team or unit because there are many definitions. Klaus and Glaser (1968) stated that "A team is usually well organized, highly structured, and has relatively formal operating procedures--as exemplified by a baseball team, an aircraft crew, or a ship control team." They further define teams as having (1) relatively rigid structure, organization, and communication networks; (2) well-defined member assignments; (3) coordinated participation of an unspecified number of individuals with specialized skills who must perform at some minimum level of proficiency; and (4) frequent involvement with equipment or tasks requiring perceptual-motor activity.

An alternative list of minimum characteristics required for a team includes (Hall & Rizzo, 1975): (1) goal- or mission-orientation, (2) formal structure, (3) assigned roles, and (4) interactions between members. The important point is to define a team in its operational context by suggesting team standards that foster achievement along the lines embedded in the context in which the team operates.

These definitions can fit a team of two individuals or multiple units that are related laterally to other units as well as vertically in a command hierarchy (Wagner et al., 1977). The development of training standards for unit performance depends on the degree of interaction and integration with other units required of the unit. If effectiveness can be improved or degraded by the performance of other units, the measuring device must be sensitive enough to determine the cause of mission success or failure. To create an evaluation system that incorporates standards with these qualities requires conducting a functional analysis of mission responsibilities from the highest level of command. The individual unit's responsibility for the accomplishment of the mission can then be determined within the context of the overall organization.

Evaluation Conditions

After the team has been defined, the next issue is to describe the situation or conditions under which the team performs. Many dimensions characterize the situation or conditions:

1. The threat faced by the team. As the threat increases, the decisions, tactics, and employment of friendly assets must respond accordingly. Evaluation of team performance in a high threat environment may involve assessing more complex interactions than an evaluation of the same task under low threat conditions.

2. The distinction between established and emergent situations. In the established situation, team responsibilities are highly structured with formal operating and communication procedures. Tasks are performed in a routine and predictable way that simplifies monitoring and evaluating the quality and timeliness. Established task situations can be viewed more simply because function assignments and responsibility among team members can be described. Evaluation criteria can be developed that list task requirements and associated proper responses.

In the emergent situation, unexpected events and little previous explicit planning for team behavior make evaluating team performance much more difficult or even impossible. Many performance solutions to an emergent problem can only be evaluated in terms of the relative success or failure of that particular event. In emergent task situations, the predictability of inputs is low. Although task assignments are defined, more than one correct response is often possible. The individual must cope with the varying environmental demands. Team members must have adaptive innovation, problem-solving, and decision-making skills. Evaluating unit performance in emergent situations demands that standards be established to measure the quality of those skills within a variety of performance situations.

3. The frequency with which they occur. They may be classified as (a) discrete tasks that occur once during an engagement, are evaluated as either performed or not performed, and are often referred to as "go/no go" tasks or (b) tasks that occur repeatedly throughout one or more phases of the mission (Wheaton, Johnson, & Dondero, 1981), which can often be evaluated in terms of subminimal, minimal, or optimal performance. Therefore, we need to develop sampling strategies that deal both with tasks that occur only once as well as with recurrent tasks.

Organizational Issues

Developing objective measures of collective performance is difficult because people perform in a complex world. The best way to measure a skill requires controlling the environment so that only the collective skill of interest occurs only at a predetermined time and place. In the real world, most skills are not performed under optimal conditions for evaluation. This is particularly true for interactive, team-based skills that require several people to perform at some minimal level of proficiency for the event to progress successfully (Wagner et al., 1977). Therefore, a context must be established to provide a basis from which to identify the criticality, conditions, and proficiency levels of unit tasks. Establishing such a context requires a method for analyzing the functions of the organization, which will provide a framework for identifying and ranking critical team tasks.

A considerable amount of research has examined how to develop models to structure, understand, and predict the outcomes of complex systems. The development of these models relies on a "top-down" function analysis of an organization that emphasizes the causal relationships between concepts, rules, and procedures within tasks and between hierarchical elements (Smith & Reigeluth, 1982). Because this effort concerns the organizational model, we will examine the implications of developing a framework to understand the functional responsibilities within a command structure.

The most important reason for developing a model is to delineate mission responsibilities at their appropriate levels. To perform a functional analysis of a Marine Corps unit, the size and complexity of the unit, the number and types of assets, and the mission responsibilities--the conceptual framework--must be determined. To establish the rules by which units perform, requires identification of the command structure, communication networks, asset tasking, and personnel allocation. Next, the procedures used to accomplish tasks that support the overall mission need to be established. Finally, responsibility for specific collective tasks can be identified in the context of the overall organization.

Analyzing the organization and its composite unit structures in terms of functional responsibility has several advantages. First, understanding the roles played by units within command echelons provides a means for predicting the interunit relationships in accomplishing tasks under specific event conditions. Unit performance requirements may differ under varying degrees of threat, mission difficulty, and situational conditions. Systematically altering those variables allows measurement of performance under a wide range of unit missions. In addition, because the unit performance standards are developed within the context of the organization, postevent analysis of individual, unit, and command echelon performance can be summarized in a significant and standardized manner.

The key organizational problem, therefore, is to develop a model to structure and predict the outcomes of unit behavior. This model will allow training officers to determine the personnel (individually or collectively) accountable for performing mission task elements, which is not possible with existing evaluation systems.

Training Issues

Individual Proficiency

Individual proficiency is necessary for effective unit performance (Kanarick, Alden, & Daniels, 1971). Unit training progresses faster when the individuals have already mastered the requisite task skills. Emphasizing team coordination too early in training may interfere with the acquisition of individual task competence (Horrocks, Krug, & Heerman, 1960; Horrocks, Heerman, & Krug, 1961). Team performance did not change when trained unit members were replaced with equally competent individuals who had not been trained with the team. Horrocks et al., (1960, 1961) agreed with Briggs and Johnson (1967) that, according to the research reviewed, no generalized team skill is independent of individual proficiencies.

As all of these task situations are established situations, unit performance seems to be the sum of individual performances. However, in an emergent situation, Johnson (1981) showed that unit training was more than the sum of the individual proficiencies when the task requires direct coordination between individuals. He concluded that unit training is more effective when the training stresses the acquisition of coordinated skills and when all possible contingencies for the tasks being trained cannot be stated and the unit must develop procedures for task accomplishment.

Performance Feedback

Feedback on the quality of team performance following an exercise event is unquestionably the single most critical parameter in team or individual training (Kanarick et al., 1971; Wagner et al., 1977). This finding is not surprising, since knowledge of results (KOR) is central to modern learning theory. Nebeker, Dockstader, and Vickers (1975) showed that units perform better with feedback than without, regardless of whether the feedback was for the individual, the unit, or both.

KOR may come from either an intrinsic source inside the individual or from an extrinsic source outside the individual. A source external to the system provides extrinsic feedback when a unit has achieved its objective and its members know that they have conducted a successful mission. Intrinsic feedback is more difficult to define because it is largely a subjective, internalized experience. Intrinsic feedback is inherent in the task itself, as when properly trained members are aware during the performance of their tasks whether they are interacting correctly or incorrectly before any extrinsic feedback is available. Further research is required to determine if either type of feedback is more effective in improving performance and, therefore, should be included in the CTS development method.

Team Training

Team training in the Navy consists of training teams organized for the performance of a particular mission (Davis, Hayes, Abolfathi, & Harvey, 1977; Wagner et al., 1977). Training is broken down into five categories: (1) preteam indoctrination or individual instruction that emphasizes increasing individual skill levels; (2) subsystem team training that consists of assigning team members to the combat systems, unit operations, or engineering systems department; (3) system subteam training that involves training two or more subsystem teams and generally an entire ship; (4) system level operational training or training at sea; and (5) multi-unit system operational training that consists of shore-based unit training before the units get underway for the exercise. These in-port exercises consist mainly of training in the tactical advanced combat direction and electronic warfare (TACDEW) system trainer.

Thurmond (1980) described a similar set of levels for the Army's team training. The instructional systems development (ISD) process for determining instructional scope and sequence involves conducting a learning analysis and identifying relationships between or among objectives. Because of the complexity of team training, the combat scenario format precludes the presentation of a precise objectives hierarchy based on the characteristics and relationships of individual tasks. The variety of team interactions present in a continuum of tactical combat situations require developing a scope and sequence for scenario presentations. Thurmond further reported that team training occurs at four levels: (1) Individual training, which assures that a minimum level of individual competence is achieved before team training can be effective, (2) beginning team training, which is doctrine training and focuses on the established team roles; (3) integrated team training, which is designed to incorporate instructional strategies that are related to coordination and compensatory member interactions; and (4) emergent team training, which should incorporate all instructional strategies previously employed as well as any new operational fluctuations and operational catastrophes identified in the job/task flow charts.

Problems with Team Performance Standards

One dominant finding in the literature is the significant lack of standards that are objective, recordable, discriminatory, and acceptable to most persons familiar with the tasks and skills of concern (Wagner et al., 1977). Standards used in evaluation systems were found to be inaccurate in some instances and too general in others.

Poorly defined standards require individual raters to exercise their own judgement to a great extent (Hayes & Wallis, 1974). In assessing the effectiveness of the Army Training and Evaluation Program (ARTEP) for evaluating unit performance, Hayes and Wallis (1974) found that the performance standards contained many indefinite terms such as on time, excessive, sufficient, proper, etc. These standards resulted in nonstandardized evaluations that significantly reduced reliability and value to the units and the training community in general. Using ARTEP also did not distinguish between tasks, conditions, measures, and standards (Wheaton et al., 1981). Consequently, they found ARTEP evaluations are characterized by unsystematic, idiosyncratic, and highly subjective judgments about unit performance. Performance of the same task often varied dramatically because of differing conditions (e.g., weather, darkness, supply conditions, etc.); hence, it is necessary to specify the measures that determine minimal standards of performance.

The Air Force assessed a measurement system used to evaluate combat-ready crew performance and found that, even in a highly sophisticated, semi-automated system, the resulting estimates of combat effectiveness were purely subjective (Obermayer, 1974). In determining the effectiveness of computer-assisted performance evaluation for the Navy's anti-air warfare training, Chesler (1971) stated that opinions about the standards for many combat situations often differed widely. He also stated that, even though objective criteria that are acceptable indices of good and poor team performance are difficult to define, they are essential for a valid assessment of unit combat readiness.

Many tasks have complex requirements; for example, many computer-based weapon systems require the coordinated actions of multiple operators and decision makers (Thurmond & Kribs 1978). These computerized command/control systems are operated by teams whose interactions with each other and the environment are mediated by the computer complex and its associated input/output requirements. These operators want observable and measurable unit training requirements that they can use to deal with these complex systems in the field. An evaluator who observes the team during an exercise evaluates team performance subjectively. Under this system, determining the effect of unit behavior for specific tasks performed by the team on final mission completion is not possible. Exercise coordinators make these decisions subjectively after the exercise.

Another problem has to do with evaluator objectivity. Wagner et al. (1977) state that the ability of human evaluators to judge unit performance in complex field exercises is unsatisfactory. Current problems experienced with field ratings are largely due to the many factors that influence the outcome of the exercise. The shortage of objective, reliable, and quantitative methods for application in field simulations is caused by the inherent difficulty of the measurement of complex, interactive human performance. Some unit evaluation methods attempt to circumvent the inherent unreliability of observers by automating detection and measurement as much as possible. Automating the monitoring and data recording processes, as with the Multiple Integrated Laser Engagement System (MILES), would circumvent the problem of unreliable, subjective human observers; however, estimating the performance of individuals and teams raises the problem that realism in training and combat are never the same. The difficulty is how to motivate teams to perform optimally and maintain the highest levels of performance in the field. Objective measurement devices that allow realistic appraisals of casualties and costs in equipment in field exercises enhance simulation fidelity, which encourages realistic performance in the field.

Some guidance for reducing ambiguous field performance evaluation have been suggested. Hall and Rizzo (1975) list the following four elements required to conduct an analysis of a unit: (1) The unit mission requirement, which is in effect the objective stating what the unit under study is to produce or achieve; (2) a measure of effectiveness, which is an objective index or scale used to determine the level of production or output of the unit; (3) a measure of the cost of the system to compute what resources need to be expended to operate the system at any level of efficiency; and (4) a combination of these inputs to yield a criterion for the final judgment of the unit. This procedure relates the effectiveness of the system to the cost of its operation.

Wagner et al. suggested that, although a systematic and applicable method for team training evaluation did not exist, the requirements of such a method could be stated as follows (1977, p. 18):

> 1. The definition of team performance objectives in terms of specified, observable outcomes to include criteria for acceptance and conditions of performance.
>
> 2. The definition of a metric or range of values applicable to each specified observable event.
>
> 3. The detection, measurement, and recording of the value of an observed event at each occurrence.
>
> 4. An evaluation of the team as having attained or not attained the objective based on the discrepancies between outcome criteria and observed event values.
>
> 5. The feedback of team performance data to the training environment.

Generally, these criteria address the concerns cited by many researchers, evaluators, and users of team performance evaluation systems. They state the requirements for developing objective, observable, and recordable performance data and allow the behaviors necessary for basic proficiency on any task to be listed. The accuracy of this checklist of behaviors will depend on the adequacy of the task definitions.

Examples of Methodologies

Research efforts have sought to develop a method for systematically reducing the subjectivity of performance standards and increasing their validity. A means for empirically evaluating the validity of the information from subject matter experts (SMEs) is needed or the subjectivity will remain in the identification of tasks to be judged, the standards and conditions of performance, the scoring methods, and the assessment of unit combat effectiveness. Several major efforts that employed different methods to develop unit performance measures are described below.

Analysis of Combat Scenarios

The Defense Advanced Research Projects Agency (DARPA) attempted to determine the critical factors in unit performance. This research, done by CACI, Inc.-Federal (Hayes et al., 1977) for DARPA, explored the feasibility of obtaining valid judgements from experienced combat officers using controlled combat situation scenarios to establish a context for their judgments. The scenarios were developed from battles fought in World War II, Korea, Viet Nam, and in specialized operations. The method was developed as an empirical measure of combat effectiveness--a scale for judging unit performance that is coherent and can be replicated. Each battle was analyzed for critical factors (i.e., quality of information, quality of plan, logistics support, awareness of enemy capabilities, maneuver during action, communication, etc.). Officers were asked to judge the performance of each unit in battle based on (1) whether the unit accomplished the mission, (2) how well the unit accomplished the mission in comparison with other units, and (3) crucial factors in the unit's success or failure. A sophisticated factor analysis of the officers' judgments provided a predictive algorithm that identified the factors associated with mission accomplishment and mission failure in terms of critical aspects of performance. These became guidelines for identifying critical tasks to be judged and developing standards and conditions under which a unit's performance should be evaluated.

Although a functional analysis of mission requirements could lead to the identification of useful data, the Marine Corps may have many future requirements that were simply not addressed during World War II, Korea, or Viet Nam. This work does, however, clearly define the need to establish performance criteria to judge combat adaptability, planning effectiveness, appropriate use of assets, and implementation of the principles of war during combat within the unit/team context.

DELPHI Technique

One well-known method for systematically extracting subject matter expertise is the DELPHI technique (Dalkey, 1968). In this technique, a small monitor team designs a questionnaire that is sent to a larger respondent group of SMEs. The monitor team summarizes the questionnaire results and bases a new questionnaire for the respondent group on these results. The respondent group usually has several opportunities to reevaluate its original answers based on the examination of the group response.

Larson, Sander, and Steinemann (1974) explored using the DELPHI technique with the Marine Corps Tactical Warfare Simulation and Evaluation system (TWSEAS). They reviewed potentially useful performance evaluation methodologies and cited research by Dalkey (1968), Helmer (1967), and Beach (1972) as providing DELPHI related methodologies applicable to measurement aspects of TWSEAS.

Larson and Sander (1975) used the DELPHI technique to identify those characteristics of unit performance that distinguish combat-ready units from noncombat-ready units as observed in battalion level field exercise environments. Larson and Sander mailed questionnaires to infantry battalion commanding officers that contained questions on (1) aspects of combat effectiveness, (2) time needed to evaluate that effectiveness, and (3) effect of incomplete information on evaluator decisions about a unit's combat effectiveness. Responses from these experts were compiled and formated to serve as feedback to individual experts for the second round of questionnaires. The same experts were then asked to rank each of the 30 to 50 items generated from their responses in round one, for each of the areas, in terms of importance and frequency. Means and standard deviations derived for expert rankings of each of these items provided a way of achieving group consensus on the relative importance of each item. Successive rounds eliminated some items and reduced the standard deviations on those selected as important factors in the three original global areas.

Advantages of the DELPHI procedure are that the SMEs can review comments made by other experienced officers, take into account the perspectives of the other officers, respond in a detailed and uninterrupted manner, and express and review stated opinions in a noncompetitive environment. This technique eliminates a substantial amount of the bias that personality differences, strength of verbal ability, and effect of rank generate in group discussions. The Larson and Sander (1975) effort yielded performance items, time requirements, and contextual factors that were later validated in the field using selected TWSEAS exercise scenarios.

An alternate approach would be to use tasks and standards derived from field exercises instead of SME opinion. With TWSEAS, the Marine Corps is in the unique position to be able to develop an evaluation system based on quasi-experimental designs that compare field performance data from different units completing identical exercises or from the same unit repeating an exercise. This method would consist of a highly structured phase in which a base-line data set for an exercise using a single, replicative scenario and a generalizing phase in which the restrictions on the scenarios are relaxed to resemble "real world" engagements (Hayes et al., 1977).

Existing Tactical Doctrine

Another approach is to use existing Marine Corps doctrine as a basis for developing CTSs. Lewellyn (1984) has approached the methodological problems with mixed results. The approach used to develop CTSs for the Marine Corps Combat Readiness Evaluation System (MCCRES) is built around mission performance standards (MPSs) that consist of tasks essential to the performance of a particular mission, the conditions under which the tasks are performed, and the requirements for the successful completion of the task. Lewellyn concludes that the standards of performance contained in the tasks and requirements of MCCRES MPS appear to be a reasonable starting point for developing CTSs. Lewellyn suggests that the process used to develop MCCRES is similar to the Army systems approach to training except for the entries relating to unit performance standards (U.S. Army Training and Doctrine Command, 1984). Initial development of MCCRES MPSs began with research on the basic mission statements formally approved and published by HQMC (Marine Corps Order 3501.2, 1977). MPS development was limited to the specific operational missions most pertinent to combat readiness. Other primary considerations for MPS development included tables of organization, current threat, techniques and doctrine, tactics, probability of actual use, and other known contingencies (Lewellyn, 1984).

Engagement Simulation

In the past, the Army based training standards on Army tactical doctrine (Lewellyn, 1984), which lists several tasks required to accomplish a given mission. A group of SMEs received a list of these tasks to refine first into critical and noncritical tasks and then divide into task steps. Under exercise conditions, the task is considered accomplished when all the steps are completed. The rating scale is usually on a go/no-go standard, although there is no fixed rule for scoring. Unit performance measurement is based on subjective standards that show how well the team works together; for example, "Each fire team works in a wedge formation."

More recently, the Army has begun an effort to select a method for the specifying unit performance variables and measures (Wheaton et al., 1981). This method uses SME opinion in a highly contextualized mission event.

The purpose of this work was to explore generalizable methods for specifying unit performance standards using an engagement simulation (ES) approach. The battlefield environment was simulated on a three-dimensional terrain board. Experienced tactical officers were selected to perform as mission planners and tacticians. Each officer was tasked to determine best course of action at each point in the terrain board game, explained his decision, and stated how that move related to or followed from his interpretation of doctrine. Before the game progressed, the officers were asked to merge or consolidate their individual plan into one consensus-based move. A group discussion was used to determine the one consolidated move. The game progressed in this fashion until the mission scenario was completed. Upon conclusion of the actual gaming session, the officers were asked to specify performance standards for each decision point (move) in the mission scenario. The players were to base their estimates on insights acquired during the course of the gaming exercise. In general, Wheaton gives three reasons for using engagement simulations (ESs): (1) to determine how well the unit adheres to and performs according to doctrine, (2) to evaluate the results of each engagement using attrition-based measures, and (3) to assess whether the finished product achieved the objective and achieved it properly. These variables and measures characterize the results of combat (Wheaton et al., 1981).

The following guidelines for the development of an ES-based evaluation system were established. The system should: (1) be superior to ARTEP in that it must also contain CTSs; (2) be driven by objective, quantitative data; (3) evaluate both processes (i.e., unit behavior and intermediate outcomes) and product (i.e., mission outcome); and (4) have criterion-referenced standards for processes, intermediate outcomes and mission outcomes. Wheaton et al. also summarized the main methodological requirements for developing such an evaluation system: a systematic definition and specification of the performance variables and measures, the objective standards for judging performance, the procedures for comparing observed performance to the performance standards, and the procedures for providing feedback to the units.

Event-based Contingency Tables

In an effort to standardize procedures for developing team training standards, Slough and Stern (1981) developed very detailed training objectives for single-ship ASW training exercises by integrating the following seven sources of information: interviews with experienced instructors, review of official Navy publications, analysis of exercise scenarios, review of grading sheets, observation of exercises and classroom lectures, and tape recordings of communications during search-attack unit (SAU) exercises.

The development procedure Slough and Stern used required: (1) observing exercises and collecting information from publications and instructors, (2) developing a contingency table relating exercise events to required team actions, (3) having the instructors review the table, (4) using the contingency table to develop the objectives, and (5) having the instructors review the objectives. Development of the objectives was based on the contingency table (Step 4). Step 4 specified in detail the team actions and the conditions leading to team actions and identified performance standards for those team actions. These standards were extremely difficult to identify and sometimes could not be identified until after the instructor review. The instructors reviewed the resultant objectives independently. Because of the required complexity of the team's performance, opinions differed on how accurately or timely an activity had to be performed. When these values were not fixed by doctrine, Slough and Stern solicited several expert opinions to determine an allowable value range or an optimum performance level.

Task Flow Analysis

Thurmond and Kribs (1978) recognized the impact of situational factors in developing a team job/task analysis technique. They established a standard operating procedure (SOP) task flow for a variety of fire mission tasks performed by a battalion-level fire-director control team (TACFIRE). Next, they interviewed experienced team members to discuss the emergent situational factors that affected the SOP and errors that led to malfunctions in the team operations. They noted that no team operates in a purely established or purely emergent situation. Therefore, the job/task and training analysis emphasized (1) defining the precise TACFIRE established situation that the SOP prescribed and (2) identifying the most common and critical emergent situations that affect operations of the TACFIRE system. By defining both the established and the emergent situations, the team member interactions could be analyzed and the teams tasks defined.

Group Consensus

The most common method for developing standards of performance for both individual and team skill requirements relies, first, on stated doctrine requirements for required task performance and, second, on a group of SMEs to rank the tasks and create the standards of performance (Lewellyn, 1984). Generally, selection of SMEs is based on their background experience and present job; that is, instructors, senior enlisted/officers, and training department personnel. The group normally has a stated period of time in which to reach a consensus on tasks and how to evaluate the performance of these tasks.

This method of developing performance standards has advantages. Associated time and personnel costs are reduced and a product is developed by the end of the meeting. Having several SMEs involved means the decisions will be more objective than if one individual made the decisions.

However, as anyone who has participated in such an effort knows, one or two highly verbal or highly ranked individuals can influence the outcome greatly. Because consensus must be achieved in a short period of time, a structured approach is often established early in the meeting. Also, because the common experience within the group in the ways the tasks were trained and evaluated, the outcome generally mirrors both the positive and negative aspects of existing training. Discussion of any real complexity in training or measuring the task tends to be discouraged because it becomes too difficult to deal with in a short period of time. Hence, the outcome tends to be simplistic in comparison to the real world requirements for individual and team performance. This over-simplification creates problems when the outcome is applied to real world events.

Validation

Each method discussed here includes procedures for deriving and applying the particular performance evaluation system in question to ensure its applicability and validity. For example, Wheaton et al. (1981) suggest the following six basic steps in an ARTEP engagement simulation validation: (1) Select the mission on which the unit will be evaluated; (2) monitor, measure, and record unit performance during the ES; (3) record the mission outcomes that the unit achieves; (4) compare the unit's behavior and outcomes to the evaluation standards; (5) give the results to unit training personnel and responsible individuals; and (6) use these results to direct later training and evaluation activities. Although these overly simplified steps assume that many complex methodological issues have been resolved, they provide a useful organizational framework to approach the problem of developing and validating a unit evaluation system.

The best choice for validating CTSs would be if an existing instrumented range could be used to develop empirically based objectives at a low cost. Even if cost were not a factor, team performance would have to be observed under a range of repeated and controlled conditions, which would probably be best on an instrumented range where performance could be monitored with a minimum of variance from evaluators and conditions. This method, if selected, would require a well defined and systematic process to collect performance data across a wide range of situations and conditions. Data collection would require substantial time, effort, and coordination between the personnel at the range and the unit standards development team. The logistics and costs of such an effort would have to be carefully considered before a commitment to this method is made.

An instrumented range might be preferable to use for validating a sample of performance standards developed using more subjective methods such as the DELPHI. The costs could be substantially reduced while providing a general indication of the validity of the newly developed unit performance standards. In short, objective (i.e., quantifiable) unit performance standards could be derived from a combination of the techniques employing SMEs and selected validations using engagement simulation.

The complexities of designing an evaluation system capable of measuring unit performance under combat conditions are readily apparent. The measures must be systematically geared to measure the quality of performance across a range of threat, difficulty, and situational conditions. They must be designed ultimately for application from squad level through Marine Air-Ground Task Force (MAGTAF) level and must integrate all available force assets. The resulting model of combat unit performance requirements weighted by conditions will provide the basis for the development of valid training standards.

The procedures used for validation of CTSs were discussed for the Army and Navy. These steps provide a useful approach to the problem of validating CTSs.

CONCLUSIONS AND DISCUSSION

From our discussion of training issues, we conclude that:

1. Effective unit performance must be built on some minimal level of individual proficiency.

2. As the task situation becomes more complex and involved (e.g., requiring original and imaginative responses to new situations), unit training can result in team skill levels that are higher than the individual skill levels.

3. Units perform better as a function of the frequency and quality of performance feedback (Nebeker et al., 1975).

The major problem cited in CTS research is developing a set of criterion variables that are objective, recordable, and discriminate between levels of performance. The inherent weakness in these standards, however, will remain if the standards are developed subjectively without an organizational context. The apparent solution to this problem is to use SMEs to analyze the unit in context to determine which tasks will be evaluated, the conditions under which the unit should be evaluated, the scoring methods, and how to assess the total effect of training. Then, this information would be subjected to an empirical validation scheme, such as the use of instrumented range, to validate the SME opinion.

Developing an evaluation system capable of producing valid measures of unit performance requires a functional analysis of the organizational responsibilities and a context for assigning mission/task responsibilities by echelon. The following conclusions assume that an organizational model defining these responsibilities will be developed and that a combination of doctrinal guidance and SMEs would accomplish the analysis. Selection of the SMEs given this responsibility would be based on their Marine Corps experience and their understanding of a broad range of mission responsibilities. SMEs selected for determining performance standards for tasks performed by specific units should be highly qualified in the technical skills and have a good understanding of the unit's interactions with the command echelons immediately above and below their unit.

The review of methodologies currently used to establish collective training/performance standards yielded a range from quasi-empirical to highly subjective development techniques. The front-end costs of all the techniques were in the same range with the quasi-empirical being the most expensive and the subjective techniques being the least expensive. However, the relative confidence in the validity of standards and quality of performance measurement is highest when derived by quasi-empirical techniques and lowest when derived by subjective techniques. The dilemma is to determine the cost/benefits inherent in selecting a particular method.

The DARPA (CACI, Inc.) effort yielded the most empirical basis for determining critical factors in unit performance However, relying on the method used in the CACI, Inc.-Federal research has drawbacks: (1) It is expensive and time consuming; (2) there was difficulty in obtaining data across all analysis variables for many candidate battle scenarios because the data were not gathered during the battle itself; (3) it forces the development of standards for future training/assessment requirements based on battles fought with weapons systems and personnel allocations up to 40 years old; and (4) there are simply not enough examples, particularly of missions that failed, to give anything more than subjective impressions of the causes of the success or failure of any particular mission.

The engagement simulation, task flow analysis, and event-based contingency table approaches all appeared to be effective in identifying team based performance standards. All three methods established an organizational context (i.e., an explicit event) and required SMEs to arrive at a consensus of opinion about specific performance requirements. However, these methods also required that the group of SMEs be assembled, and particularly in the engagement simulation method, reach a group-based consensus of opinion. In addition, these efforts required an on-site facilitator staff to direct, monitor, and record SME decisions. Although the methods are effective, their utilization is costly in terms of manpower scheduling and facilitating. That cost will reduce the number of training events that can be addressed in short periods of time.

SMEs making decisions in a group, which is the least expensive solution, is not an optimal method for developing performance objectives because of the inherent pressures of personalities and time constraints. In particular, this would be the poorest method for determining the unit performance standards within the context of the overall organizational structure. Historically, groups of SMEs tend to focus too narrowly on one unit and to neglect the interunit and command structure for the tasks of interest.

The DELPHI technique would provide an expedient compromise between the high validity and relatively high cost of reconstructing combat scenarios and the low validity and low cost of a group-based SME task analysis. Outcomes of previous efforts using the DELPHI method have been favorably received by evaluation system users, training personnel, and instructor staffs. Clearly, tactical doctrine requirements would guide the selection of the tasks to be evaluated and provide any other pertinent guidance. SMEs could participate in an iterative development process that would allow the perspectives of other experts to surface and be considered without the pressures of time and personality differences.

DELPHI procedures avoid the limitations of group procedures, which include excess influence of highly verbal or highly ranked individuals, shortage of time to discuss complicated and involved issues, and the tendency for the outcome to be simplistic when compared to real world requirements for CTS.

RECOMMENDATIONS

1. Develop a method for determining unit functional responsibilities within mission tasks. The resulting data would serve as a basis for determining collective training requirements and their supporting individual training requirements.

2. Develop a method for identifying and formatting effective standards for guiding the training and evaluation of collective training requirements.

3. Investigate the feasibility of the use of the DELPHI approach in obtaining a consensus of SME opinion for the establishment of collective training standards.

4. Develop cost-effective approaches for validating the effectiveness of collective training standards.

REFERENCES

Beach, L. R. (1972). Studies in the psychology of decisions. Seattle, WA: University of Washington.

Briggs, G. E., & Johnson, W. A. (1967). Team training (Final Report NAVTRADEVCEN-1327-4). Orlando, FL: Naval Training Device Center.

Chesler, D. J. (1971). Computer-assisted performance evaluation for Navy anti-air warfare training: Concepts, methods and constraints (Research Report SRR 71-25). San Diego: Navy Personnel and Training Research Laboratory.

Dalkey, N. C. (1968). Predicting the future. Santa Monica, CA: Rand Corporation,

Defense Science Board. (1975). Summary report of the task force on training technology. Washington, DC: Office of the Director of Defense Research and Engineering, Department of Defense.

Glaser, R., & Klaus, D. (1962). Proficiency measurement: Assessing human performance. Robert Gagne (Ed.), In Psychological Principles in System Development. New York: Holt, Rinehart & Winston.

Hall, E. R., & Rizzo, W. A. (1975). An assessment of U.S. Navy tactical team training (TAEG Report No. 18). Orlando, FL: Training Analysis and Evaluation Group.

Hayes, R. E., & Wallis. (1974). ARTEP Validation Report. Alexandria, VA: U.S. Army Research Institute for Behavioral and Social Sciences.

Hayes, R. E., Davis, P. C., Hayes, J. J., Abolfathi, F., & Harvey, B. (1977). Measurement of unit effectiveness in Marine Corps infantry battalions (Technical Report CAC3810002AG). Arlington, VA: CACI, Inc.-Federal.

Headquarters, United Stated Marine Corps. (9 December 1977). Marine Corps Combat Readiness Evaluation System (MCCRES), Vol. I, Introduction (Marine Corps Order 3501.2). Washington, DC: Author.

Helmer, O. (1967). Analysis of the future: The DELPHI method (P-3558). Santa Monica, CA: Rand Corporation.

Horrocks, J. E., Krug, R. E., & Heermann, E. (1960). Team training II: Individual learning and team performance (Technical Report NAVTRADEVCEN 198-2). Port Washington, NY: Naval Training Device Center.

Horrocks, J., Heermann, E., & Krug, R. E. (1961). Team training III: An approach to optimum methods and procedures (Technical Report NAVTRADEVCEN 198-3). Port Washington, NY: Naval Training Device Center.

Johnson, E., Wheaton., J. R., Allen, T. W., Forrester, R. E., & Sulzen, R. H. (1981). Research on systems analysis approaches to unit evaluation: Specification of performance standards for ES-based ARTEP (Tech. Rep.). Washington, DC: American Institutes for Research.

Kanarick, A. F., Alden, D. G., & Daniels, R. W. (1971). Decision making and team training in complex tactical training systems of the future. In 25th Anniversary Commemorative Technical Journal, Naval Training Device Center.

Larson, O. A., & Sander, S. I. (1975). Development of unit performance effectiveness measures using DELPHI procedures (NPRDC Tech. Rep. 76-12). San Diego: Navy Personnel Research and Development Center. (AD-A015 963)

Larson, O. A., Sander, S. I., & Steinemann, J. H. (1974). Survey of unit performance effectiveness measures (NPRDC Tech. Rep. 74-11). San Diego: Navy Personnel Research and Development Center. (AD-774 919)

Lewellyn, M. T. (1984). Collective training standards analysis (CNA Working Paper 84-1245.09). Alexandria, VA: Center for Naval Analyses.

Nebeker, D. M., Dockstader, S. L., & Vickers, R. R. (1975). A comparison of the effects of individual and team performance feedback upon subsequent performance (NPRDC Tech. Rep. 75-35). San Diego: Navy Personnel Research and Development Center. (AD-A010 131)

Obermayer, R. W. (1974). Combat ready crew performance measurement system (AFHRL Final Rep. TR-74-108(I)). Air Force Human Resources Laboratory.

Rocklyn, E. H., Jacobs, R. R., Magy, M. A., & Archibald, A. G. (1975). A method for increasing the training effectiveness of Marine Corps tactical exercises: A pilot study (NPRDC Tech. Rep. 75-34). San Diego: Navy Personnel Research and Development Center. (AD-A013 224)

Slough, D. A., & Stern, H. W. (1981). Development of antisubmarine warfare team training objectives (NPRDC Tech. Note 81-18). San Diego: Navy Personnel Research and Development Center.

Smith, J. P., & Reigeluth, C. M. (1982). The structural strategies model (NPRDC Tech. Note 82-18). San Diego: Navy Personnel Research and Development Center.

Thurmond, P., & Kribs, H. D. (1978). Computerized collective training for teams: Final report (ARI Tech. Rep. 78-A1). Alexandria, VA: U.S. Army Research Institute for the Behavioral & Social Sciences.

Thurmond, P. (1980). Development of analysis, design and development techniques for team ISD (Tech. Rep. PR001-80-005). San Diego: Path Research.

U.S. Army Training and Doctrine Command. (1984). System approach to training (TRADOC Reg 350-7), Coordination Draft.

Wagner, H., Hibbits, N., Rosenblatz, R., & Schulz, R. (1977). Team training and evaluation strategies: A state-of-the-art review (HumRRO Tech. Rep. SR-ED-76-11). Alexandria, VA: Human Resources Research Organization.

Wheaton, G. R., Johnson, E., & Dondero, L. J. (1981). Research on system analytic approaches to unit evaluation: Specification of performance variables and measures for ES-based ARTEP (Tech. Rep.) Washington, DC: American Institutes for Research.

DISTRIBUTION LIST

Chief of Naval Operations (OP-11), (OP-11H), (OP-987H)
Chief of Naval Research (Code 270)
Chief of Naval Education and Training (Code N-21)
Director, Naval Education and Training Program Development Center
Commanding Officer, Naval Education and Training Program (Code 09) (Personnel and Training Research)
Naval Education and Training Support Center, Pacific
Chief of Naval Technical Training (Code 00)
Commanding Officer, Naval Training Equipment Center (Technical Library) (5), (Code 10), (Code 7)
Commandant of the Marine Corps (MPI-20)
Mr. Greenup, Education Center MCDEC
Commander, Army Research Institute for the Behavioral and Social Sciences, Alexandria (PERI-POT-I), (PERI-SZ)
Commander, Air Force Human Resources Laboratory, Brooks Air Force Base (TSRL/Technical Library FL 2870)
Commander, Air Force Human Resources Laboratory, Lowry Air Force Base (AFHRL/ID)
Superintendent, Naval Postgraduate School
Director of Research, U.S. Naval Academy
Institute for Defense Analyses, Science and Technology Division
Center for Naval Analyses
Canadian Forces Personnel, Applied Research Unit, Canada
DPSRSC, National Defense Headquarters, Canada
Ministry of Defense, Senior Psychologist, England (Naval)
D. Dennison, Army Personnel Research Establishment, Personnel Psychological Division, England (2)
Science 3 (RAF), Lacon House, England
1 Psychological Research Unit, NBH 3-44, Australia
Directorate of Psychology - AF, Department of Defense (Air for CE), Australia
Navy Psychology, Australia (2)
Defense Psychology Unit, Defense HQ, New Zealand (2)
Dr. Regan
E. Ramras
Defense Technical Information Center (DDAC) (2)